Calculating State Dependent Noise in a Linear Inverse Model Framework

Cristian Martinez-Villalobos*

Department of Atmospheric and Oceanic Sciences, University of Wisconsin – Madison, Madison, Wisconsin, and Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California

Daniel J. Vimont

Department of Atmospheric and Oceanic Sciences, University of Wisconsin – Madison, Madison, Wisconsin

Cécile Penland

Physical Sciences Division, NOAA/Earth System Research Laboratory, Boulder, Colorado

Matthew Newman

Cooperative Institute for Research in Environmental Sciences, University of Colorado, and NOAA/ESRL, Boulder, Colorado

J. David Neelin

Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California

[*]Corresponding author address: Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, 520 Portola Plaza Math Sciences Building 7127 Los Angeles CA 90095, U.S.A.

E-mail: cmartinezvil@atmos.ucla.edu

ABSTRACT

The most commonly used version of a Linear Inverse Model (LIM) is forced by state independent noise. Although having several desirable qualities, this formulation can only generate long-term Gaussian statistics. LIM-like systems forced by correlated additive-multiplicative (CAM) noise have been shown to generate deviations from Gaussianity, but parameter estimation methods are only known in the univariate case, limiting their use for the study of coupled variability. In this paper we present a methodology to calculate the parameters of the simplest multivariate LIM extension that can generate long-term deviations from Gaussianity. This model (CAM-LIM) consists of a linear deterministic part forced by a diagonal CAM noise formulation, plus an independent additive noise term. This allows the possibility of representing asymmetric distributions with heavier-or lighter-than-Gaussian tails. The usefulness of this methodology is illustrated in a locally coupled 2 variable ocean-atmosphere model of midlatitude variability. Here a CAM-LIM is calculated from Ocean Weather Station data. Although the time resolved dynamics is very close to linear at a time scale of a couple of days, significant deviations from Gaussianity are found. In particular individual probability density functions are skewed with both heavy and light tails. It is shown that these deviations from Gaussianity are well accounted for by the CAM-LIM formulation, without invoking nonlinearity in the time resolved operator. Estimation methods using knowledge of the CAM-LIM statistical constraints provide robust estimation of the parameters with data lengths typical of geophysical time series, e.g., 31 winters for the Ocean Weather Station here.

1. Introduction

Multivariate linear theory has been used to great success in practically all realms of climatic science. One widely applied linear method is the Linear Inverse Model (LIM) (Penland and Sardeshmukh 1995) framework, in which a linear approximation to a system dynamics is empirically obtained from the a system's covariance statistics. In this framework, a linearly stable system describing the evolution of a "slow" variable anomalies (e.g. sea surface temperatures anomalies), is driven by Gaussian white noise representing the effect of unresolved "fast" variability (e.g. wind stress, convection, etc) on the slow variable (Papanicolaou and Kohler 1974; Penland 1996). It is a common practice to restrict the noise forcing the LIM to be state independent (additive), and while often providing valuable results, it is not required by these kinds of systems. This kind of model has been used successfully as a forecast tool (Newman 2013), and performs well when the underlying slow deterministic dynamics is linear or weakly non-linear.

Despite the qualitative (and often quantitative) success of linear inverse models, these kind of models are unable in general to reproduce observed deviations from Gaussianity, when driven by additive Gaussian white noise. These deviations from Gaussianity are typified for example in skewed (asymmetric) or kurtotic (lighter or heavier than Gaussian distribution tails) probability density functions (PDFs). Deviations from Gaussianity in geophysical variables distributions are commonplace and well documented (e.g., Monahan 2004; Neelin et al. 2010; Ruff and Neelin 2012; Stefanova et al. 2013; Loikith et al. 2013; Perron and Sura 2013; Cavanaugh and Shen 2014; Huybers et al. 2014; Loikith and Neelin 2015; Sardeshmukh et al. 2015), and can be generated through multiple dynamical processes. Perhaps the most intuitive of these mechanisms is through nonlinearity in the deterministic dynamics, with the models of Timmermann et al. 2001, Kratsov et al. 2005, Kondrashov et al. 2006, Chen et al. 2016 (among others), providing examples in the inverse modeling setting. Simple advective-diffusive prototypes for passive tracers under a mean gradient can produce distinct non-Gaussianity, most evidently at the distribution tails (Bourlioux and Majda 2002; Neelin et al. 2010). Other mechanisms that lead to non-Gaussianity include cross frequency coupling (Rennert and Wallace 2009), jet stream meandering (Luxford and Woollings 2012), first passage processes (Stechmann and Neelin 2014; Neelin et al. 2017). Sura and Hannachi (2015) provide a comprehensive review on the mechanisms that generate deviations from Gaussianity in the atmospheric sciences.

Alternatively, even if the deterministic term (i.e., the term in which noise is not explicit) is linear, deviations from Gaussianity may arise through interactions between a slowly evolving system and fast transients forcing the system, if the fast transients depend on the state of the system (Sura et al. 2005). Strictly speaking, any differential equation with stochasticity in it represents a treatment of nonlinearity at some level. That is where dynamical stochasticity originates. A linear system forced with additive noise represents a coarse-graining long enough that all of the state dependence, if any, of the nonlinear effects is averaged out. In that case, the Central Limit Theorem (CLT) applies strongly enough to render the statistics of the measured state approximately Gaussian. When the timescale separation between the linear decay and the rapid non-linearities is too small to invoke such a strong version of the CLT, but is large enough to average out the details of the nonlinearities, the system may be modeled as a linear process with state dependent (multiplicative) noise. Thus, unlike additive noise, the multiplicative noise processes that drive the deterministic dynamics explicitly depend on the system state (e.g. sub-daily wind variance dependence on storminess or blocking, or surface fluxes depending on local stability). Multiplicative noise is well established as a source of non-Gaussianity (Penland 2003; Sura et al. 2005; Majda et al. 2008; Sardeshmukh and Sura 2009; Franzke et al. 2015; Sura and Hannachi 2015; Berner et al. 2017), and has been employed to model several aspects of climate variability including El

Niño Southern Oscillation (Perez et al. 2005; Jin et al. 2007; Levine and Jin 2017), and extra tropical variability (Neelin and Weng 1999; Sura et al. 2005).

For evaluation and comparison purposes, it is important to establish a baseline for variability, including deviations from Gaussianity, that can be explained through a multilinear deterministic system that integrates (possibly) state dependent noise. In order to do that it is necessary to have a simple methodology to extract the multiplicative noise information from data. This has proven difficult because the state dependent noise, as elaborated below, in general contributes to both the "signal" and the "noise", so disentangling its contribution is not straightforward. Thus, despite important progress on the matter (e.g., Siegert et al., 1998; Peavoy et al., 2015), a simple methodology to calculate the state dependent noise from data in a statistically consistent way has been lacking. The development of this methodology, tailored to linear deterministic systems driven by multiplicative noise, is the primary goal of this paper.

In general, fast variability may depend not only on the magnitude of the system anomalies, but also on their sign. This to a first approximation can be modeled through a type of noise formulation termed Correlated Additive-Multiplicative (Müller 1987; Sura et al. 2006; Sardeshmukh and Sura 2009; Majda et al. 2009; Penland and Sardeshmukh 2012; Sardeshmukh and Penland 2015; Sardeshmukh et al. 2015; Franzke 2017) noise or CAM noise. Mathematically the CAM noise amplitude depends linearly on the state of the system, and this dependency is allowed to be asymmetric with respect to the mean. This asymmetry is expected in systems where the fast variability is modulated differently whether the system is in its positive or negative state, which naturally leads to skewness. This is the case when linearizing the effects of rapid wind variability on fluxes affecting ocean mixed layer dynamics (Sura et al. 2006; Sura and Newman 2008). For example Sura et al. 2006, studying an ocean mixed layer model, finds at least two (related) sources for this noise amplitude asymmetry. The first one arises due to ocean-atmosphere mean state temperature differences. This affects the sensible and latent heat fluxes driven by rapid wind variability at the ocean-atmosphere interface, and can be mapped directly onto a CAM noise term. The second source arises due to the different sensitivity of boundary layer stability to positive or negative anomalies. This contribution, while not precisely following a CAM noise form (a piecewise linear function would be better), can be approximated by it.

In addition to the noise amplitude asymmetry, the CAM noise linear state dependency is important because it modifies the probability of noise events as the system evolves, leading to higher probability of extreme events (at least in one tail), compared to similar systems forced by pure additive noise. In fact, in the univariate case it can be shown that the skewness S and excess kurtosis $K-3$ are related such that[1] (Sura and Sardeshmukh 2008; Sardeshmukh and Sura 2009)

$$K-3 \geq \frac{3}{2}S^2. \tag{1}$$

Several variables have been found to follow such a parabolic $K-3 \geq \frac{3}{2}S^2 - \delta$ relationship (Sardeshmukh and Sura 2009; Sardeshmukh and Penland 2015; Sardeshmukh et al. 2015; Sura and Hannachi 2015), where $\delta > 0$ is a small offset that occurs possibly due to sampling effects. In other words, this framework produces heavy tailed distributions (although considering the skewness generated one of the tails may be light at values less than about 10 standard deviations. At larger values, the tails behave similarly. We ignore these extreme tails in what follows.), and is an attractive candidate to correctly model extreme events (Sardeshmukh et al. 2015).

Henceforth in this paper, we will consider the next step in complexity beyond estimating parameters from the standard LIM (driven by additive noise) and LIM applied to the univariate CAM system (Sardeshmukh et al. 2015). That is, we consider a Linear Inverse Model driven by a simplified diagonal CAM noise formulation (CAM-LIM). Although this formulation neglects CAM

[1] Note that Sardeshmukh et al. 2015 derives a stricter bound $K-3 \geq \frac{15}{8}S^2$. This is discussed in section 3b in the context of the multivariate system presented here.

noise covariance and nonlocal state dependency (see for example Sardeshmukh and Sura 2009 equations 4a-4b), it is a more general model than used in previous applications, and allows for the generation of deviations from Gaussianity in a linear deterministic setting.

To calculate CAM noise in a LIM setting, consistency relations between the CAM-LIM parameters and the statistics generated by it will be derived. In this way a statistical dynamical description of a system is calculated, which can be employed for multiple purposes, including the construction of realistic forecasts and representation of its scatter, as well as the study of the underlying processes that generated the observations. Importantly the employment of this model can be used as a baseline for the variability expected from deterministic linear dynamics, and raises the bar for claims of nonlinear behavior. In order to do this we use the Stratonovich Fokker-Planck equation (Fokker 1914; Kolmogoroff 1931. See Gardiner 2010 for a discussion of Ito (Ito 1951) and Stratonovich (Stratonovich 1966) calculi) ;

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\sum_i \frac{\partial (A_i(\mathbf{x},t)p(\mathbf{x},t))}{\partial x_i} - \frac{1}{2}\sum_{i,j,m} \frac{\partial}{\partial x_i}(\frac{\partial F_{im}(\mathbf{x},t)}{\partial x_j} F_{jm}(\mathbf{x},t)p(\mathbf{x},t))$$
$$+\frac{1}{2}\sum_{i,j,m} \frac{\partial^2}{\partial x_i x_j}(F_{im}(\mathbf{x},t)F_{jm}(\mathbf{x},t)p(\mathbf{x},t)), \tag{2}$$

which is the equation satisfied by the PDF of a deterministic system driven by Gaussian white noise:

$$\frac{dx_i}{dt} = A_i(\mathbf{x},t) + \sum_m F_{im}(\mathbf{x},t)\eta_m. \tag{3}$$

In this equation A_i encodes the deterministic dynamics and F_{im} the amplitude of noise process η_m affecting variable x_i, and Ito's circle is implied. For future reference we will clarify the terminology used in (2). The first term in that equation corresponds to the "deterministic drift", the second term is known as the "noise induced drift" and is zero if the noise is independent of the state of the system, and the last term is usually called the "diffusion". For a heuristic explanation of the noise induced drift see Sura and Newman 2008 (section 2). It is worth pointing out that in the LIM framework, only a combination of deterministic drift and noise induced drift, known as effective drift, can be inferred from data, rather than the terms separately (Penland 2007). An important result from the framework presented herein is that, within the confines of this model (section 3), the deterministic and noise induced drifts can be separately resolved.

Stochastic modeling has been used to study different aspects of climate variability (see Berner et al. 2017 for a review). In particular simplified versions of (3) have provided important insight into the nature of ocean-atmosphere interactions in the mid latitudes (e.g., Frankignoul and Hasselmann 1977, Hall and Manabe 1997, Barsugli and Battisti 1998, Sura et al. 2006, Sura and Newman 2008). We will illustrate the derivation of the CAM-LIM parameters, and the general usefulness of the model by constructing a 2 variable model of ocean-atmosphere thermal coupling in mid latitudes, empirically derived from an Ocean Weather Station data. The remainder of this manuscript is organized as follows. Section 2 presents a brief overview of the LIM framework. Section 3 introduces the CAM-LIM, some important simplifications, and the derivation of the parameters of the model as a function of its statistical structure. Additionally the constraint (1) is updated to include the effects of the coupling. Section 4 exemplifies this in the previously mentioned 2 variable thermal coupling model, and results are compared to the standard LIM modeling of the same system. Finally, section 5 concludes the paper.

2. Brief Review of Linear Inverse Modeling

In this section we present a brief overview of the LIM (Penland and Sardeshmukh 1995). In this framework an N component state vector of anomalies $\mathbf{x}$ evolve according to the following linear equation (also written in component notation for future use):

$$\frac{d\mathbf{x}}{dt} = \mathbf{M}\mathbf{x} + \mathbf{S}\eta \tag{4}$$

$$\frac{dx_i}{dt} = \sum_{j=1}^{N} M_{ij}x_j + \sum_{l=1}^{L} S_{il}\eta_l. \tag{5}$$

In this equation $\mathbf{M}$ is a constant $N \times N$ matrix, $\mathbf{S}$ is a state independent $N \times L$ matrix of noise amplitudes, and η is a L component vector of Gaussian white noise processes. Note that the noise covariance matrix $\mathbf{S}\mathbf{S}^{\mathbf{T}}$ has an $N \times N$ dimensionality. The matrix $\mathbf{M}$ denotes the slow time resolved linearized dynamics, while the temporally unresolved fast variability is modeled by the noise input $\mathbf{S}\eta$. In this framework $\mathbf{M}$ is a stable operator so the system needs the stochastic input to generate variance. Here the diagonal terms ($M_{ii} < 0$) correspond to an effective measure of dissipating processes that depend linearly on variable x_i and the system is coupled through the M_{ij} $(i \neq j)$ terms. Finally, the matrix $\mathbf{M}$ can be calculated from data (von Storch et al. 1988; Penland and Sardeshmukh 1995) using:

$$\mathbf{M} = \frac{1}{\tau} log(\mathbf{C}_\tau \mathbf{C}_\mathbf{0}^{-\mathbf{1}}). \tag{6}$$

where $\mathbf{C}_\tau =< \mathbf{x}(\tau)\mathbf{x}(0)^T >$ is the lag covariance matrix at lag τ, and $\mathbf{C}_0 =< \mathbf{x}(0)\mathbf{x}(0)^T >$ is the contemporaneous covariance matrix. Here $<>$ denotes a long term average.

Given an initial condition $\mathbf{x}(0)$, the most probable evolution $\mathbf{x}(t)$ of the system is (Penland 2007)

$$\mathbf{x}(t) = e^{\mathbf{M}t}\mathbf{x}(0). \tag{7}$$

There is one key difference in how this multilinear system behaves compared to its univariate version ($x(t) = e^{-\lambda t}x(0)$, $\lambda > 0$). In absence of stochastic forcing the one-dimensional system decays exponentially, while in the multilinear case short-term growth is possible if the dynamics of the system are non-normal ($\mathbf{M}\mathbf{M}^{\mathbf{T}} \neq \mathbf{M}^{\mathbf{T}}\mathbf{M}$, e.g. Boyd 1983, Farrell 1988, Borges and Hartmann 1992, Penland and Sardeshmukh 1995, Moore and Kleeman 1999, Thompson and Battisti 2000, Zanna and Tziperman 2005, Vimont 2010, Sévellec and Fedorov 2017, Martinez-Villalobos and Vimont 2017). This makes possible the use of this framework as a forecasting tool (Penland and Sardeshmukh 1995; Penland 1996; Johnson et al. 2000; Alexander et al. 2008; Newman et al. 2011; Zanna 2012).

There are balance conditions in the dynamics of stochastically generated systems that can be deduced from the Fokker-Planck equation (2). In statistical steady state, the Fluctuation-Dissipation relation (e.g. Leith 1975, Penland and Matrosova 1994 , DelSole and Hou 1999, Ghil et al. 2002, Gritsun et al. 2008) relates the state variables covariance $\mathbf{C_0} =< \mathbf{x}\mathbf{x}^T >$ to the noise processes covariance $\mathbf{SS}^T$ as:

$$\mathbf{MC_0} + \mathbf{C_0M^T} + \mathbf{SS}^T = 0 \tag{8}$$

where we also write this relation in component notation for future reference

$$\sum_l (M_{nl} < x_l x_k > + < x_n x_l > M_{kl}) + \sum_m S_{nm} S_{km} = 0. \tag{9}$$

This can be understood as a covariance budget, where the fluctuating stochastic input is dissipated by the deterministic dynamics, so statistical steady state is attained.

The LIM framework is and has been used extensively to study the state of the tropical Pacific (Penland and Sardeshmukh 1995; Penland 1996; Newman et al. 2011; Vimont et al. 2014; Capotondi and Sardeshmukh 2015), tropical Atlantic (Penland and Matrosova 1998; Vimont 2012), as well as extra tropical dynamics (Alexander et al. 2008; Zanna 2012; Newman 2013; Newman et al. 2016). In the tropical Pacific the forecast of sea surface temperature (SST) anomalies through this method is competitive compared to forecasts provided by General Circulation Models (Newman and Sardeshmukh 2017). The LIM framework provides a good description of the state variables contemporaneous and lagged covariances if the temporally resolved dynamics is close to linear, but it is not designed to account for long term deviations from Gaussianity, for example asymmetric behavior between positive and negative anomalies, and different than Gaussian frequency of extreme events.

3. Linear Inverse Model driven by Correlated Additive-Multiplicative noise (CAM-LIM)

In this section we introduce a CAM-LIM framework, calculate several formulas to extract the multiplicative noise information from data, and derive and discuss the constraints that this formulation puts on the statistical moments generated.

a. Model Derivation

In order to retain the advantages of the LIM approach, and also account for deviations from Gaussianity while keeping the modifications to a minimum, we consider a LIM-type model driven by a simple CAM noise formulation, assuming diagonal dominance in the multiplicative term. Similarly to the standard LIM, a slow variable integrates fast random forcing, but in this case the random forcing amplitude depends on the slow variable state itself. The model is given as follows:

$$\frac{dx_i}{dt} = \sum_{j}^{N} A_{ij}x_j + \sum_{m=1}^{N} (G_i + E_i x_i)\delta_{im}\eta_m + \sum_{m=N+1}^{L} B_{im}\eta_m - D_i \tag{10}$$

Here x_i corresponds to the i component of a state vector $\mathbf{x}$ of anomalies and $\mathbf{A}$ is an $N \times N$ matrix that encodes the linearized deterministic dynamics of the system. Entries $A_{ii} < 0$ corresponds to deterministic dissipating processes that depend linearly on x_i, and the system is coupled through the A_{ij} terms ($i \neq j$). The system is driven by L Gaussian white noise processes η_m whose amplitudes F_{im} (in keeping with the notation of equation 3) are given as follows

$$F_{im} = (G_i + E_i x_i)\delta_{im} \quad \text{for } m = 1 \text{ to } N$$

$$F_{im} = B_{im} \quad \text{for } m = N+1 \text{ to } L. \tag{11}$$

The first set of coefficients ($(G_i + E_i x_i)\delta_{im}$) correspond to the CAM noise processes. Here $E_i x_i$ corresponds to a "local" state dependency for the noise amplitude, and G_i accounts for the part of the additive noise that is correlated to the state dependent (multiplicative) noise. The second set of coefficients (B_{im}) denote the amplitude of additive noise processes uncorrelated to the CAM noise. For simplicity this formulation neglects direct nonlocal noise state dependency, although part of the nonlocal effects can be captured (if local and nonlocal variables are correlated) through this simple local state dependency. In this formulation the CAM noise processes affect the individual noise variances (as seen below), while the pure additive noise carries the noise covariances information. An important feature of this model is that the noise amplitude is asymmetric with respect to the mean, i.e. the magnitude of the CAM noise amplitude is zero at $x_i = -\frac{G_i}{E_i}$ rather than at $x_i = 0$. This will produce an expected mean noise induced drift that can be removed from the equation for the anomalies (10) by a term $D_i = \frac{1}{2}E_i G_i$ (Sardeshmukh and Sura 2009). In the univariate case this model corresponds exactly to the one proposed and solved by Sardeshmukh and Sura 2009.

The use of a diagonal CAM noise formulation (one independent process per variable) and the neglect of direct nonlocal noise state dependency are important simplifications, but allows us to calculate relatively simple formulas for the CAM-LIM parameters. Using this particular CAM noise formulation is the logical first step to introduce noise state dependency in a LIM framework, and it is in the spirit of, though more general than, the principle of diagonal dominance postulated by Sardeshmukh and Sura (2009, section 6). This principle states the increasing importance of the self correlation terms in representing the higher order statistics of a system, and explains the success of the univariate version of this model in representing the observed deviations from Gaussianity in several climate variables (Sardeshmukh and Sura 2009; Penland and Sardeshmukh 2012; Sardeshmukh et al. 2015; Sura and Hannachi 2015). Here in addition to the terms considered by

Sardeshmukh and Sura, coupling between the variables and noise covariance effects are incorporated. This allows for the calculation of joint statistics. Despite these simplifications, in most cases the model will be enough to display a realistic representation of the emergent non-Gaussian behavior, while maintaining all the advantages of the standard LIM framework.

Multiplying the Fokker-Planck equation (2) by the appropriate moment of $\mathbf{x}$ and integrating over from -∞ to ∞ we can calculate an equation for the first two moments of the system. In statistical steady state,

$$\frac{d < x_k >}{dt} = \sum_l (A_{kl} + \frac{1}{2} E_k^2 \delta_{kl}) < x_l >= 0 \tag{12}$$

$$\begin{aligned} \frac{d < x_n x_k >}{dt} &= \sum_l ((A_{nl} + \frac{1}{2} E_n^2 \delta_{nl}) < x_l x_k > + < x_n x_l > (A_{kl} + \frac{1}{2} E_k^2 \delta_{kl})) \\ &+ \sum_m B_{nm} B_{km} + G_n^2 \delta_{nk} + E_n^2 < x_n^2 > \delta_{nk} = 0 \end{aligned} \tag{13}$$

Comparing to (9) and imposing that both standard LIM and CAM-LIM describe the first two moments of the system in the same way, the following relations obtain:

$$M_{kl} = A_{kl} + \frac{1}{2} E_k^2 \delta_{kl}. \tag{14}$$

$$(SS^T)_{nk} = (BB^T)_{nk} + G_n^2 \delta_{nk} + E_n^2 < x_n^2 > \delta_{nk} \tag{15}$$

These relations relate the parameters of a standard LIM to the parameters of a CAM-LIM. Here (14) makes explicit the partition of the effective dissipating processes $M_{ii} x_i$ into a deterministic part $A_{ii} x_i$, and a noise induced modification $\frac{1}{2} E_i^2 x_i$. Also (15) enforces that both standard LIM and CAM-LIM reproduces the same noise covariance, with the right hand side of the expression amounting to a partition of it between pure additive terms and CAM noise processes. Formulas to calculate all these terms from data are derived in the appendix, with some important ones repeated below.

Under CAM-LIM, it can be shown that the best prediction (in the mean square sense) of the evolution of the state vector [2] given a current state $\mathbf{x}(0)$ is also given by (7) (Penland 2007)

$$\mathbf{x}(t) = e^{\mathbf{M}t}\mathbf{x}(0) \tag{16}$$

which further justifies the use of the notation shown in (14). Also (14) and (16) reiterate the message that in general when calculating the matrix $\mathbf{M}$ from data, that determination not only includes the linearized deterministic drift, but also a noise-induced drift component that may be confused with deterministic dynamics (Penland and Matrosova 1994). Equation (13) generalizes the fluctuation-dissipation relation to include the extra CAM noise terms. From the Fokker-Planck equation we can also calculate an equation for the system (unnormalized) skewness ($< x_k^3 >$) and kurtosis ($< x_k^4 >$) budgets. Again, in statistical steady state:

$$\frac{d < x_k^3 >}{dt} = 3\sum_l M_{kl} < x_l x_k^2 > +6E_k G_k < x_k^2 > +3E_k^2 < x_k^3 >= 0 \tag{17}$$

$$\frac{d < x_k^4 >}{dt} = 4\sum_l M_{kl} < x_l x_k^3 > +6(\sum_m B_{km}^2 + G_k^2) < x_k^2 > +12E_k G_k < x_k^3 > +6E_k^2 < x_k^4 >= 0. \tag{18}$$

Combining the information provided by the first four statistical moments (equations 12, 13, 17, 18) we may find an expression for the CAM-LIM parameters as

$$E_j^2 = \frac{(-2\overline{K}_{jj} + 3\overline{S}_{jj}S_{jj} + 6\overline{V}_{jj})}{3(K_{jj} - 1 - S_{jj}^2)} \tag{19}$$

$$G_j = -\frac{1}{2}\frac{C_{jj}^{1/2}}{E_j}(E_j^2 S_{jj} + \overline{S}_{jj}) \tag{20}$$

$$(BB^T)_{jj} = -(2\overline{V}_{jj} + E_j^2)C_{jj} - G_j^2 \tag{21}$$

where matrices $\mathbf{V}$, $\mathbf{S}$, and $\mathbf{K}$ entries are defined as

$$V_{ij} = \frac{< x_i x_j >}{< x_j^2 >} \equiv \frac{C_{ij}}{C_{jj}} \qquad S_{ij} = \frac{< x_i x_j^2 >}{< x_j^2 >^{3/2}} \qquad K_{ij} = \frac{< x_i x_j^3 >}{< x_j^2 >^2}, \tag{22}$$

[2] In this case the mean of the conditional PDF will not correspond in general to its most probable value (Penland 2007).

matrices denoted with a bar are defined as

$$\overline{\mathbf{V}} = \mathbf{MV} \qquad \overline{\mathbf{S}} = \mathbf{MS} \qquad \overline{\mathbf{K}} = \mathbf{MK}, \tag{23}$$

and C_{ij} denote particular entries of the covariance matrix $\mathbf{C_0}$ (C_{jj} is the variance of variable x_j). The non-diagonal elements of $\mathbf{BB^T}$ are calculated using (15). Notice that S_{jj} and K_{jj} are just the skewness and kurtosis of variable x_j. Note that in the multivariate case shown here variable x_l influences x_k ($l \neq k$) skewness and kurtosis through $\mathbf{M}$. Analogous to the univariate case (Sardeshmukh and Sura 2009) the statistics generated by the CAM-LIM are constrained in a distinctive way. These constraints are explored in more detail in the section below. Remaining aspects of the derivation are shown in the appendix.

b. CAM-LIM constraints on the statistics

In general, the moments of a CAM-LIM generated dataset (10) are necessarily constrained. The first constraint (denoted as C_1) can be derived from (19) and is given as follows[3] for variable x_j

$$C_1(x_j) = -\overline{K}_{jj} + \frac{3}{2}\overline{S}_{jj}S_{jj} + 3\overline{V}_{jj} \geq 0. \tag{24}$$

This constraint reduces to (1) in the univariate case (which is a good consistency check), and shows that given a non-zero real amplitude of the multiplicative noise term, the CAM-LIM will generate variability that is typically kurtotic. This is a manifestation of the increased chances for the system to make extreme event excursions, due to the noise amplitude state dependency.

A second constraint arises because the pure additive covariance matrix $\mathbf{BB}^T$ needs to be positive definite (see equation 15). This constraint (denoted as C_2) may be written as

$$C_2 = det(\mathbf{BB}^T) \geq 0; \tag{25}$$

[3]Notice that the denominator of (19) is always positive (Wilkins 1944).

This constraint necessarily, but not sufficiently, requires the following inequality (denoted with a $'$) to be satisfied as well (equation 21)

$$C_2'(x_j) = -(2\overline{V}_{jj} + E_j^2)C_{jj} - G_j^2 > 0. \tag{26}$$

The last inequality, given that C_1 has already been satisfied, ensures that the additive noise variances are positive. Basically this limits the contribution of the CAM noise to the total noise covariance. In the univariate case simultaneous consideration of constraints C_1 and C_2 leads to an stricter relation between kurtosis (K) and skewness (S) $K - 3 \geq \frac{15}{8}S^2$ (Sardeshmukh et al. 2015). Although a similar (but more complicated) relation could be derived in the multivariate case, here we keep both constraints separate. These relations will be explored in practice in section 4c.

4. Modeling mid latitude ocean-atmosphere local coupling using CAM-LIM

In this section we apply the CAM-LIM methodology to a simple dataset that has been investigated in the literature (Hall and Manabe 1997; Sura et al. 2006; Sura and Newman 2008). A simple model of ocean-atmosphere coupling in the mid latitudes is calculated from data, and compared to observations. The CAM-LIM parameters estimation procedure is described in detail, and the information provided by the constraints described above is used to improve the calculation of the parameters.

a. The Models

Simple linear stochastic models have been extensively used to study ocean-atmosphere interactions (e.g., Frankignoul and Hasselmann 1977; North and Calahan 1981; Kim and North 1992; Hall and Manabe 1997; Barsugli and Battisti 1998; Sura et al. 2006; Wu et al. 2006; Sura and Newman 2008; Smirnov et al. 2014). These kinds of systems are simple enough that can be regarded as a null hypothesis or baseline against which distinctively nonlinear variability can be compared. Here we show the usefulness of this framework by modeling the local midlatitude ocean-atmosphere coupling using both standard LIMs and CAM-LIM frameworks. The CAM-LIM and standard LIM are given as follows

$$\frac{dT_i}{dt} = \sum_j A_{ij} T_j + \sum_{l>2} B_{il}\eta_l + (G_i + E_i T_i)\eta_i - \frac{1}{2} E_i G_i \qquad \textbf{CAM-LIM} \tag{27}$$

$$\frac{dT_i}{dt} = \sum_j M_{ij} T_j + \sum_l S_{il}\eta_l \qquad \textbf{LIM} \tag{28}$$

where T_i is the i component ($i = 1,2$) of vector $\mathbf{T} = [T_a \; T_o]^T$. Here T_a and T_o represent near surface atmospheric and surface oceanic temperature anomalies at a particular mid latitude location. Standard LIM and CAM-LIM parameters are defined as in (5) and (10) respectively, and can be calculated using (6) and (9) in the standard LIM case, and (6), (19), (20), and (21) in the CAM-LIM case. LIM and CAM-LIM parameters are related as in (14), and (15). Although nonlocal noise state dependency (i.e., $\frac{dx_i}{dt} = ... + E_{ij} x_j \eta$ terms, $i \neq j$) is expected for this kind of interaction (e.g. Neelin and Weng 1999, Sura and Newman 2008), the simple CAM noise formulation used here provides satisfactory results (as seen below), especially compared to a standard LIM. Interestingly, within the confines of this model formulation, the noise part $\frac{1}{2}E_i^2$ and deterministic part A_{ii} contributions to T_i effective damping term M_{ii} can be cleanly separated out using this framework. Below we show the result of the previously stated calculations.

b. Models Parameter Estimation

To estimate parameters for our models (27) and (28) we use Ocean Weather Station (OWS) data (For information on OWS see Diaz et al. 1987, and Dinsmore 1996), specifically OWS Papa (OWS P) in the North Pacific. OWS P is located far from strong currents (Hall and Manabe 1997), and is only affected weakly by ENSO (Alexander et al. 2002), thus providing an ideal location to construct these models.

We consider daily data from January 1 1950 to December 31 1980 (total 31 years). $\bar{T}_a$ and $\bar{T}_o$ climatologies are constructed using the annual mean plus the first three annual Fourier harmonics. Anomalies (T_a and T_o) are computed by subtracting the respective daily climatologies. The few unavailable daily values ($\sim 1.5\%$ of the total) are neglected when computing the climatologies, and February 29 values are neglected as well. A 3 day running mean is applied to the anomalies, and only "extended winter" (November to April) values are considered to construct the model. Finally T_a and T_o are standardized for easier comparison. Note that using standardized variables is only done for further plotting convenience. To help gauge the results the standard deviations are $\sigma(T_a) = 1.30^oC$ and $\sigma(T_o) = 0.67^oC$.

The parameter estimation algorithm starts with the calculation of $\mathbf{M}$ from data using (6). This requires $\mathbf{T} = [T_a\ T_o]^T$ contemporaneous and lag covariance matrices. For our calculations we use a lag τ of 6 days. Notice that both LIM and CAM-LIM generate the same lag covariance matrix as required by (7) and (16). Importantly, both linear models provide an excellent representation of the observed lag correlation functions, as seen in figure S1 (Supplemental Material). The remaining model parameters are calculated using (9) for the standard LIM case, and (19), (20), (21), (13), and (14) for the CAM-LIM case. The sensitivity of the E_i and G_i calculated values to the choice of lag is fairly minor, with maximum variations respect to the values quoted below of the order of $\sim 10\%$ for reasonable choices of lag (figure S2). The results for the CAM-LIM model are given as follows:

$$\mathbf{M} = \begin{pmatrix} -0.231 & 0.069 \\ 0.013 & -0.025 \end{pmatrix} \qquad E_1 = 0.139 \;\; E_2 = 0.046 \;\; G_1 = -0.397 \;\; G_2 = 0.087$$

$$\mathbf{A} = \begin{pmatrix} -0.241 & 0.069 \\ 0.013 & -0.026 \end{pmatrix} \qquad \mathbf{BB^T} = \begin{pmatrix} 0.222 & 0.037 \\ 0.037 & 0.028 \end{pmatrix} \qquad \mathbf{C_0} = \begin{pmatrix} 1 & 0.462 \\ 0.462 & 1 \end{pmatrix}. \tag{29}$$

We notice that the effect of the state dependent noise on the damping of each variable is relatively minor (compare A_{11} with M_{11} for example). The values of E_i (the amplitude of the multiplicative noise) and G_i (the amplitude of the additive noise correlated to the multiplicative noise) differ from what would be calculated in an univariate setting (uncoupled system, no noise covariance). For example E_1 and E_2 would be overestimated by 12% and 27% [calculated using equation 19 in the univariate case ($M_{ij} = 0$ when $i \neq j$), or alternatively using Sardeshmukh et al. 2015 equation 8] had we assumed individual, CAM noise driven, univariate models for T_a and T_o.

It is tempting to compare the calculation of these parameters (29) to Sura and Newman (2008) modeling of the same dataset (their equations 29, 34, 36). Although superficially similar, the two models differ in several respects making the comparison difficult. The model presented here is totally empirical, while Sura and Newman's takes into account the dynamical equations. Having somewhat different objectives, the two models make different assumptions which prohibit their direct comparison. For example while the CAM-LIM simplified noise formulation allows for a direct estimation of the noise amplitudes, it will not directly represent some of the nonlocal effects in Sura and Newman's model. It is important to emphasize that in the CAM-LIM case there are no assumptions as to where the noise is coming from, whereas Sura and Newman neglect some potentially important processes (ocean currents, vertical entrainment, variable mixed layer depth, mixing) in order to highlight deviations from Gaussianity arising from the effect of state dependent rapid wind fluctuations on sensible and latent heat fluxes at the air-sea interface. Due to the positive mean climatological ocean-atmosphere temperature difference almost everywhere, models restricted to local air-sea interaction can only generate positive SST skewness (Sura and Sardeshmukh 2009). Although SST skewness is positive at OWS P, there are many parts of the globe where skewness is negative (Sura and Sardeshmukh 2008; Sardeshmukh and Penland 2015). Comparing to the dimensional reduction strategy employed in Sura and Sardeshmukh 2009 (their equation 16 or 19), CAM-LIM independent T_a deterministic components on (27) allows for the parameterization of other processes, besides air-sea temperature difference. This implies that unlike models restricted to local air-sea interactions, CAM-LIM is able to generate negative SST skewness as well, if the data support it. Despite these differences, the two types of models (loosely speaking "empirical" and "dynamical") are complementary and taken together help inform the relative importance of local air-sea interaction vs other processes.

To compare both the standard LIM (28) and CAM-LIM (27) with observations we run both models 10 times for 1000 years each with the calculated parameters (29) using the stochastic Heun integration method (Rümelin 1982; Ewald and Penland 2009). We remove the first 50 years of each integration as spin up time, for a total of 9500 years of LIM and CAM-LIM generated time series. We use an integration time step of 3 minutes and collect daily output. This corresponds to 9500 full years of (3 day running mean) daily values, or equivalently to 19157 extended winters of 181 days.

Using the generated datasets we calculate the T_a and T_o joint PDFs produced by each model (Fig. 1b for standard LIM, and Fig. 1c for CAM-LIM), and we compare them with the observed joint PDF in figure 1a. The joint PDFs are calculated using a bivariate Gaussian kernel density estimator, and shading denotes the difference from a best fit bivariate normal distribution. As expected the standard LIM produces a Gaussian joint PDF. On the other hand, although there are differences at the finer scale, the CAM-LIM performs noticeably better at reproducing the observed deviations from Gaussianity. Visually, some of the differences between the observed and CAM-LIM joint PDFs may look important, most strikingly what appears to be two local maxima separated by a local minimum. Here we note that similar "inhomogeneities" in the joint PDF do arise in other contexts, most notably in the study of atmospheric "regimes" (e.g., Kimoto and Ghil 1993, Smyth et al. 1999), where they are usually explained as arising through nonlinear deterministic dynamics. It is shown below that those inhomogeneities in this case likely appear due to limited sampling and are well explained by the CAM-LIM framework.

Given the extended LIM and CAM-LIM integrations one may ask how the observations compare with LIM and CAM-LIM integrations of the same length. Figure 2 shows the difference between the observed joint PDF and Monte-Carlo estimates for the LIM joint PDF (2a) and CAM-LIM joint PDF (2b). For each model Monte-Carlo PDF estimates are obtained for each of 617 different 31 year periods (181 extended winter days per year) contained within the respective 9500yr simulations, and averaged to obtain the dashed curve. Shading indicates regions where the observed PDF falls outside of the 2.5th or 97.5th percentile calculated from the 617 LIM and CAM-LIM PDF estimates. Comparing 2a and 2b it is visually apparent that the observed variability can be better explained through the CAM-LIM formulation. Although there are some spots where the observed and CAM-LIM joint PDFs are different (at the 95% confidence level), noticeably for strong positive T_a, for the most part the CAM-LIM provides a good model to explain the observed variability, including the deviations from Gaussianity. We note that both LIM and CAM-LIM have problems explaining the largest T_a anomalies, although that problem is much more reduced in the CAM-LIM case. Here we point out that the inhomogeneities in the observed joint PDF are non significant and can be well explained by a CAM-LIM null hypothesis at the 95% confidence level. In addition, only one local maxima in the observed joint PDF deviates significantly from Gaussian as seen in fig. 2a. Given the good correspondence between observed and CAM-LIM joint PDFs, it is suggested that even a coarse noise state dependency, as presented here, may significantly improve coupled variability statistics.

A similar analysis can be conducted for the distribution of the individual variables. Figure 3 shows the observed, standard LIM and CAM-LIM generated T_a and T_o cumulative density functions (CDFs) in a linear axis. Similarly as before confidence intervals are calculated using a Monte-Carlo procedure. An important difference between the standard LIM and CAM-LIM is that CAM-LIM generates asymmetric confidence intervals, –with narrower spread for positive T_a and negative T_o, where the noise amplitudes are smaller (see equation 29)–, whereas the LIM generates symmetric confidence intervals. The top panels (3a,b) show the CDFs in the middle range of the data (between -2 and 2 standard deviations). Both observed T_a (3a) and T_o (3b) CDFs are well within the 95% confidence interval generated by both LIM and CAM-LIM (not shown), although even in this range the CAM-LIM fit is noticeably better. The middle panel and lower panel shows the CDFs at the negative tails (3c,d) and positive tails (3e,f) respectively. For clarity figures 3c-f are also shown in a logarithmic y axis in figure S3. As seen in these panels, it is for extreme events where the differences between the standard LIM and CAM-LIM are most evident. With the exception of the largest positive T_a anomalies ($T_a > 2.5\sigma(T_a) \approx 3.3^oC$, see figure S3c), the CAM-LIM produces a better fit of the observed variability at the tails, including both light and heavy tails. For example this is seen in the heavier than Gaussian tail of negative T_a, and the lighter than Gaussian tail of negative T_o. With only the aforementioned exception, the observations stay within the 95% confidence level generated by the CAM-LIM realizations, whereas for the most part that is not the case for the standard LIM, where only the T_o negative tail is well captured. To put numbers in perspective, a negative T_a value of 3 standard deviations (an anomaly of $\sim -4^oC$) occurs 5 times more frequently in both observations and CAM-LIM, than in the standard LIM.

A general understanding of the data distribution, including the behavior of the tails, can be found by calculating the distribution's skewness and kurtosis. Table 1 shows the observed skewness and kurtosis, as well as the values calculated using the full LIM and CAM-LIM integrations. As expected the standard LIM skewness and kurtosis matches the ones of a Gaussian distribution. Even though the match is not perfect, it is evident that the CAM-LIM provides a closer match to observations.

There is an important degree of variability in the statistics as a function of the length of the data segment considered for the calculations. Figure 4 shows the skewness (S) and excess kurtosis ($K-3$) distributions when partitioning the standard LIM and CAM-LIM generated time series in segments of 31 winters (the length of the OWS P observations) as done before. First, note that although the fitting works better for T_a than T_o, in both cases the observed skewness and excess kurtosis are within the 95% confidence interval generated by the CAM-LIM realizations. Conversely the observed skewness and kurtosis values fall outside the standard LIM confidence interval in all cases, implying that the observed deviations from Gaussianity are a feature of this locally coupled system, and are not due to limited sampling. Second, note that the values of skewness and kurtosis in the different CAM-LIM realizations are fairly variable. For example there are several segments where T_o and T_a excess kurtosis is bigger than 2 ($K-3$ 99th percentiles are 2.21 and 3.36 respectively), implying a much higher than average number of extreme events over that interval. On the other hand, for example, there are segments where T_o excess kurtosis is negative, meaning that although the system generates long-term heavy tailed variability, quiet extreme events periods are not unusual. This variability shows that the CAM-LIM generative process (equation 27) supports a wide range of 31 years climates. This implies that for this system important swings, owing to internal dynamics, in the number of extreme events decade to decade, or even century to century, is what is normal, rather than the anomaly. This has important consequences, for example, for hypothesis testing of extreme events (Sardeshmukh et al. 2015).

c. Parameter Estimation and CAM-LIM generated statistical constraints

In this section we analyze how well the parameter calculation algorithm (19), (20), (21) performs on the CAM-LIM generated variability, that uses (29) as input parameters. This is an important self-consistency check as the output parameters from the estimation procedure should match the input parameters. When using the full CAM-LIM integration as our time series we retrieve the following values:

$$\mathbf{M} = \begin{pmatrix} -0.232 & 0.068 \\ 0.013 & -0.025 \end{pmatrix} \qquad E_1 = 0.140 \;\; E_2 = 0.046 \quad G_1 = -0.397 \;\; G_2 = 0.085$$
$$\mathbf{A} = \begin{pmatrix} -0.241 & 0.068 \\ 0.013 & -0.026 \end{pmatrix} \qquad \mathbf{BB^T} = \begin{pmatrix} 0.224 & 0.037 \\ 0.037 & 0.028 \end{pmatrix}. \qquad \mathbf{C_0} = \begin{pmatrix} 1.001 & 0.462 \\ 0.462 & 1.004 \end{pmatrix}. \tag{30}$$

The retrieved parameters compare very well with the input (29) with differences starting on the third decimal value, showing that the methodology is self-consistent (i.e., input parameters are related to the statistics generated from (19, 20, 21)). As is the case for most stochastically generated systems, a long segment of data is needed for the retrieved parameters (30) to match the input parameters (29), and there will be some inherent variability when using shorter segments of the data, as shown below.

Although the observational input data (and by construction the full CAM-LIM integration) satisfy the CAM-LIM constraints (24, 25, 26), for short enough segments of the data sampling variability may cause these constraints to be not satisfied. Practically this becomes a problem because these "short enough" segments may be longer than the available data set for a particular application. To partially overcome this we need a redefinition of the sample statistics that take into account constraints (24, 25, 26). This can be done in several ways. Taking into account the information provided by the constraints we choose a simple redefinition of matrix $\mathbf{K}$ on equation (22) (recall the entry K_{jj} corresponds to variable j kurtosis) as the rescaled matrix $(1+\alpha)\mathbf{K}$ ($\alpha \geq 0$). This is similar to the methodology used by Sardeshmukh et al. (2015) in the univariate version of this problem. The strategy is as follows. We increase α up to the point the first constraint (24) is satisfied for both variables[4]. But, as α is increased the second constraint (25) may or may not be satisfied. In most cases increasing α just to the point where the first constraint is satisfied results in E_j values that are barely above zero, implying large G_j values in order to satisfy the skewness and kurtosis budgets (17,18). If a particular G_j is too large (26) is likely not satisfied. This implies that given the statistics, E_j and G_j values are constrained to be inside a surface defined by C_1 (24) and C_2 (25,26). Basically, increasing α allows us to find the interval of values that E_j and G_j can take to stay inside that surface. Although we are not guaranteed an unbiased calculation of E_j and G_j, following this procedure, we recover values that are at least within the much narrower band of possible values.

As an example, figure 5 shows the histogram of E_1 retrieved values when using data segments that match the length of the observational input (31 winters, $31 \times 181 = 5611$ daily values, fig. 5a,c), and 100 winters ($100 \times 181 = 18100$ daily values, fig. 5b,d) of the full CAM-LIM integration ($\sim 3.47 \times 10^6$ daily values). For visual clarity there is a kernel density estimation of the distribution of E_1 values superimposed to each histogram. In each 31 or 100 winters partition we show two different cases, one denoted "$\alpha = 0$" and one denoted "α varying". The $\alpha = 0$ case shows the distribution of the retrieved values for the segments where both constraints (24, 25) are satisfied

[4]This provides a conservative estimate. In this step we may choose to calculate the constraint variable by variable, and some variables may be recovered faster.

without modification of matrix $\mathbf{K}$ (38% for the 31 winters segment length case, and 69% of the time for the 100 winters case, see Fig. 5e), and the α varying case the distribution of retrieved values after the procedure described above is followed ($\mathbf{K}$ redefined as $(1+\alpha)\mathbf{K}$; note that this includes the $\alpha = 0$ instances). The bright green point denotes the value of the input E_1 parameter calculated from the observational data (equation 29). Figure 5e shows the percentage of times where the constraints are satisfied as a function of α. Some general takeaways from this figure are as follows. As might be expected, the longer the segment considered, the better representation of the long term statistics, and the faster (24) and (25) are satisfied (Fig. 5e). Also expectedly, segments that satisfy the aforementioned constraints without modification of $\mathbf{K}$ ($\alpha = 0$) provide a better match of the long term statistics, though considerable sampling variability exists. Finally, although it can be further refined, the procedure of redefining the sample $\mathbf{K}$ matrix produces reasonable estimations, meaning that in this case approximated values can be retrieved by redefinition of the sample statistics. This result may prove useful in practice when using CAM-LIM to model other systems. Note that in general the noise terms are much harder to estimate. For example a length of 500 winters is needed for the standard deviation of E_1 retrieved values to be within 10% of the input value (29). On the other hand, as expected, the "effective drift" values are estimated much faster, for example only needing segments of ~ 25 winters for the M_{11} retrieved values standard deviation to be within 10% of the M_{11} input value.

5. Concluding Remarks

In this paper we consider a natural extension of the Linear Inverse Model framework. Here in addition to an additive Gaussian white noise component, the system is driven by a simple state dependent noise formulation, termed CAM noise (Sardeshmukh and Sura 2009). Compared to a standard LIM, this framework generates the same (lag and contemporaneous) covariance structure and the same expected evolution of anomalies, while at the same time generating skewness and excess kurtosis. One important result is that the statistical moments generated by this system are constrained. One of the constraints identified here generalizes the well known univariate CAM noise constraint (equation 1) between skewness (S) and kurtosis (K) to include the effects of coupling and noise covariance. In common with the univariate case, the coupled time series generated are typically kurtotic, making this an attractive framework to model extreme events frequencies in many cases. The univariate constraint has been shown to be relevant for different climate variables (Sardeshmukh and Sura 2009; Sardeshmukh and Penland 2015; Sardeshmukh et al. 2015). We expect the multivariate constraint (24) to provide additional information for coupled datasets.

We illustrate the general framework by using a locally coupled model of ocean-atmosphere interaction in mid latitudes. We calculated the model parameters using available sea surface temperature T_s and near surface atmospheric temperature T_a at an Ocean Weather Station. We show that, compared to a standard LIM, the CAM-LIM better reproduces the joint PDF of T_a and T_o as well as the individual PDFs. Importantly, both light and heavy tails are better described by the CAM-LIM formulation, which may be of interest also in the modeling of lighter than Gaussian tails (e.g. Loikith and Neelin 2015). Practical issues related to the implementation of the model, including the amount of data needed, were also discussed. An important point here is that knowledge of the statistical constraints arising from this framework can be used to improve the parameter estimation in cases where there is insufficient data to adequately resolve the statistics of the system.

Although here we presented the concrete example of a mid latitude coupled model, we picture this framework to have wide applicability. In specific, any system where the time resolved dynamics is reasonably linear, but significant deviations from Gaussianity are present, is susceptible to be modeled using CAM-LIM. Here we note that the model has been tested in other contexts, including higher dimensional systems, with good results (Martinez-Villalobos 2016). Implicit in the derivation of this framework is a separation of the dynamics between slow and fast timescales. Here we argue (together with many other studies) for the importance of the fast unresolved part of the dynamics in shaping not only the variance of the resolved dynamics, but also the mean state (through the noise induced drift), asymmetry in the PDF, and the behavior of the extremes. The tool presented here can be valuable to quantify these effects.

Acknowledgments. This work was supported by National Science Foundation Grant AGS-1463643 (CM, DJV and MN), and National Science Foundation Grant AGS-1540518 (CM and JDN).

APPENDIX

Appendix: Derivation of CAM-LIM parameters.

The starting point here is the Fokker-Planck equation (2) which applies to a system of the form (3). First we start by rewriting (10) as

$$\frac{dx_i}{dt} = \sum_j^N A_{ij}x_j + \sum_{m=1}^{N} B^M_{im}\eta_m + \sum_{m=N+1}^{L} B^A_{im}\eta_m - D_i \tag{A1}$$

Here $<\eta_m(t)>= 0$, $<\eta_m(t)\eta_n(t')>= \delta(t-t')\delta_{mn}$, and we have explicitly separated the CAM noise coefficients $B^M_{im} = (G_i + E_i x_i)\delta_{im}$ and pure additive noise coefficients $B^A_{im} = B_{im}$. Writing (A1) in equation (3) form we have

$$\begin{aligned}
A_i &= A_{ij}x_j - D_i \\
F_{im} &= B^M_{im} \qquad (m = 1 \textbf{ to } N) \\
F_{im} &= B^A_{im} \qquad (m = N+1 \textbf{ to } L)
\end{aligned} \tag{A2}$$

Separating the Fokker-Planck equation (2) into its deterministic drift (DD), noise-induced drift (ND), and diffusion (DI) parts, we have

$$\frac{dp}{dt} = DD + ND + DI \tag{A3}$$

Using (A2) DD, ND, and DI are given as follows;

$$DD = -\sum_{i,j}^{N} A_{ij} \frac{\partial}{\partial x_i}(x_j p) + \sum_{i}^{N} D_i \frac{\partial p}{\partial x_i} \tag{A4}$$

$$ND = -\sum_{i,j}^{N} \frac{1}{2} E_i^2 \delta_{ij} \frac{\partial}{\partial x_i}(x_j p) - \frac{1}{2}\sum_{i}^{N} E_i G_i \frac{\partial p}{\partial x_i} \tag{A5}$$

$$\begin{aligned} DI = &\frac{1}{2}\sum_{i=1}^{N}\left(G_i^2 \frac{\partial^2 p}{\partial x_i^2} + 2E_i G_i \frac{\partial^2 (x_i p)}{\partial x_i^2} + E_i^2 \frac{\partial^2 (x_i^2 p)}{\partial x_i^2}\right) \\ &+ \frac{1}{2}\sum_{i,j}^{N}\sum_{m=N+1}^{L} B_{im} B_{jm} \frac{\partial^2 p}{\partial x_i x_j} \end{aligned} \tag{A6}$$

Equations (A5), and (A6) make explicit the CAM-noise processes enter the system in both noise induced drift and diffusion parts, while the pure additive noise processes only enter in the diffusion. Equations (A4) and (A5) can be combined into an “effective” drift (ED, $ED = DD + ND$) term as

$$ED = -\sum_{i,j}^{N} M_{ij} \frac{\partial}{\partial x_i}(x_j p). \tag{A7}$$

Here $M_{ij} = A_{ij} + \frac{1}{2} E_i^2 \delta_{ij}$ as (14) and we have identified the mean noise induced drift response D_i to be equal to $\frac{1}{2} E_i G_i$. After all previous steps the Fokker-Planck equation is the addition of (A6) and (A7)

$$\begin{aligned} \frac{dp}{dt} = &-\sum_{i,j}^{N} M_{ij} \frac{\partial}{\partial x_i}(x_j p) + \frac{1}{2}\sum_{i=1}^{N}\left(G_i^2 \frac{\partial^2 p}{\partial x_i^2} + 2E_i G_i \frac{\partial^2 (x_i p)}{\partial x_i^2}\right. \\ &\left. + E_i^2 \frac{\partial^2 (x_i^2 p)}{\partial x_i^2}\right) + \frac{1}{2}\sum_{i,j}^{N}\sum_{m=N+1}^{L} B_{im} B_{jm} \frac{\partial^2 p}{\partial x_i x_j}. \end{aligned} \tag{A8}$$

Multiplying (A8) by the appropriate moment and integrating from $-\infty$ to ∞ we obtain equations (12, 13, 17, 18) in the main text. Focusing in the diagonal terms, and in statistical equilibrium, this implies the following set of equations that need to be satisfied simultaneously

$$\frac{d<x_k>}{dt} = \sum_{j=1}^{N} M_{kj} <x_j>= 0 \tag{A9}$$

$$\begin{aligned}\frac{d<x_k^2>}{dt} &= 2\sum_{j=1}^{N} M_{kj} <x_j x_k> + G_k^2 + E_k^2 <x_k^2> \\ &+ \sum_{m=N+1}^{L} (B_{km})^2 = 0\end{aligned} \tag{A10}$$

$$\begin{aligned}\frac{d<x_k^3>}{dt} &= 3\sum_{j=1}^{N} M_{kj} <x_j x_k^2> + 6E_k G_k <x_k^2> \\ &+ 3E_k^2 <x_k^3>= 0\end{aligned} \tag{A11}$$

$$\begin{aligned}\frac{d<x_k^4>}{dt} &= 4\sum_{j=1}^{N} M_{kj} <x_j x_k^3> + 6G_k^2 <x_k^2> \\ &+ 12E_k G_k <x_k^3> + 6E_k^2 <x_k^4> \\ &+ 6\sum_{m=N+1}^{L} (B_{km})^2 <x_k^2>= 0.\end{aligned} \tag{A12}$$

Here (A9) is used to eliminate the mean terms ($< x_j >= 0$). Then (A10) and (A12) are used to simultaneously eliminate G_k and B_{km}^2 terms, obtaining the expression for E_k^2 (equation 19 main text), as a function of $\mathbf{M}$ (previously calculated using (6)), and the system statistics. Here we can calculate $A_{kj} = M_{kj} - \frac{1}{2}E_k^2\delta_{kj}$. Interestingly, the skewness budget (A11) is independent of the pure additive noise amplitude. We use that information to calculate G_k (equation 20 main text) as a function of $\mathbf{M}$, E_k, and the statistics of the system. Finally $\sum_{m=n+1}^{n+l}(B_{km})^2$ (equation 21 main text) is calculated as the remainder needed to close the variance budget. It is well known that only the quadratic expression $\sum_{m=n+1}^{n+l}(B_{km})^2$ rather than the individual amplitude terms B_{km} can be extracted from data (e.g. Monahan 2004, Sura and Newman 2008). In this simplified system there are unique expressions for E_k (up to a $\pm$ sign) and G_k, but if more complex CAM noise formulations are specified, only quadratic E_k, and G_k forms will be extracted from data.

References

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The Atmospheric Bridge: The Influence of ENSO Teleconnections on AirSea Interaction over the Global Oceans. *Journal of Climate*, **15 (16)**, 2205–2231, doi:10.1175/1520-0442(2002)015⟨2205:TABTIO⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}282002{\%}29015{\%}3C2205{\%}3ATABTIO{\%}3E2.0.CO{\%}3B2.

Alexander, M. A., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear Inverse Model Predictions of the PDO. *Journal of Climate*, **21 (2)**, 385–402, doi:10.1175/2007JCLI1849.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1849.1.

Barsugli, J. J., and D. S. Battisti, 1998: The Basic Effects of AtmosphereOcean Thermal Coupling on Midlatitude Variability*. *Journal of the Atmospheric Sciences*, **55 (4)**, 477–493, doi:10.1175/1520-0469(1998)055⟨0477:TBEOAO⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281998{\%}29055{\%}3C0477{\%}3ATBEOAO{\%}3E2.0.CO{\%}3B2.

Berner, J., and Coauthors, 2017: Stochastic Parameterization: Toward a New View of Weather and Climate Models. *Bulletin of the American Meteorological Society*, **98 (3)**, 565–588, URL http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-15-00268.1.

Borges, M. D., and D. L. Hartmann, 1992: Barotropic Instability and Optimal Perturbations of Observed Nonzonal Flows. *Journal of the Atmospheric Sciences*, **49 (4)**, 335–354, doi:10.1175/1520-0469(1992)049⟨0335:BIAOPO⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281992{\%}29049{\%}3C0335{\%}3ABIAOPO{\%}3E2.0.CO{\%}3B2.

Bourlioux, A., and A. J. Majda, 2002: Elementary models with probability distribution function intermittency for passive scalars with a mean gradient. *Physics of Fluids*, **14 (2)**, 881–897, doi:10.1063/1.1430736, URL http://aip.scitation.org/doi/10.1063/1.1430736.

Bowman, A. W., and A. Azzalini, 1997: *Applied smoothing techniques for data analysis : the kernel approach with S-Plus illustrations*. Clarendon Press, Oxford :, 193 pp., URL https://searchworks.stanford.edu/view/3749403.

Boyd, J. P., 1983: The Continuous Spectrum of Linear Couette Flow with the Beta Effect. *Journal of the Atmospheric Sciences*, **40 (9)**, 2304–2308, doi:10.1175/1520-0469(1983)040⟨2304:TCSOLC⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281983{\%}29040{\%}3C2304{\%}3ATCSOLC{\%}3E2.0.CO{\%}3B2.

Capotondi, A., and P. D. Sardeshmukh, 2015: Optimal precursors of different types of ENSO events. *Geophysical Research Letters*, **42 (22)**, 9952–9960, doi:10.1002/2015GL066171, URL http://doi.wiley.com/10.1002/2015GL066171.

Cavanaugh, N. R., and S. S. P. Shen, 2014: Northern Hemisphere Climatology and Trends of Statistical Moments Documented from GHCN-Daily Surface Air Temperature Station Data from 1950 to 2010. *Journal of Climate*, **27 (14)**, 5396–5410, doi:10.1175/JCLI-D-13-00470.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-13-00470.1.

Chen, C., M. A. Cane, N. Henderson, D. E. Lee, D. Chapman, D. Kondrashov, and M. D. Chekroun, 2016: Diversity, Nonlinearity, Seasonality, and Memory Effect in ENSO Simulation and Prediction Using Empirical Model Reduction. *Journal of Climate*, **29 (5)**, 1809–1830, doi:10.1175/JCLI-D-15-0372.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0372.1.

DelSole, T., and A. Y. Hou, 1999: Empirical Stochastic Models for the Dominant Climate Statistics of a General Circulation Model. *Journal of the Atmospheric Sciences*, **56 (19)**, 3436–3456, doi:10.1175/1520-0469(1999)056⟨3436:ESMFTD⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281999{\%}29056{\%}3C3436{\%}3AESMFTD{\%}3E2.0.CO{\%}3B2.

Diaz, H. F., C. S. Ramage, S. D. Woodruff, and T. S. Parler, 1987: Climatic summaries of ocean weather stations. URL https://docs.lib.noaa.gov/noaa{_}documents/OAR/ERL{_}ARL/Diaz1987.pdf.

Dinsmore, R., 1996: Oceanus. *Oceanus*, URL http://www.whoi.edu/oceanus/feature/alpha-bravo-charlie.

Epanechnikov, V. A., 1969: Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & Its Applications*, **14 (1)**, 153–158, doi:10.1137/1114019, URL http://epubs.siam.org/doi/10.1137/1114019.

Ewald, B., and C. Penland, 2009: Numerical Generation of Stochastic Differential Equations in Climate Models. 279–306, doi:10.1016/S1570-8659(08)00206-8, URL http://linkinghub.elsevier.com/retrieve/pii/S1570865908002068.

Farrell, B., 1988: Optimal Excitation of Neutral Rossby Waves. *Journal of the Atmospheric Sciences*, **45 (2)**, 163–172, doi:10.1175/1520-0469(1988)045⟨0163:OEONRW⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281988{\%}29045{\%}3C0163{\%}3AOEONRW{\%}3E2.0.CO{\%}3B2.

Fokker, A. D., 1914: Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. *Annalen der Physik*, **348 (5)**, 810–820, doi:10.1002/andp.19143480507, URL http://doi.wiley.

com/10.1002/andp.19143480507.

Frankignoul, C., and K. Hasselmann, 1977: Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29 (4)**, 289–305, doi:10.1111/j.2153-3490.1977.tb00740.x, URL http://tellusa.net/index.php/tellusa/article/view/11362.

Franzke, C. L. E., 2017: Extremes in dynamic-stochastic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **27 (1)**, 012 101, doi:10.1063/1.4973541, URL http://aip.scitation.org/doi/10.1063/1.4973541.

Franzke, C. L. E., T. J. O'Kane, J. Berner, P. D. Williams, and V. Lucarini, 2015: Stochastic climate theory and modeling. *WIREs Climate Change*, **6**, 63–78, doi:10.1002/wcc.318, URL http://onlinelibrary.wiley.com/doi/10.1002/wcc.318/full.

Gardiner, C. W., 2010: *Stochastic Methods : A Handbook for the Natural and Social Sciences*. Springer Berlin.

Ghil, M., and Coauthors, 2002: Advanced spectral methods for climatic time series. *Reviews of Geophysics*, **40 (1)**, 1003, doi:10.1029/2000RG000092, URL http://doi.wiley.com/10.1029/2000RG000092.

Gritsun, A., G. Branstator, and A. Majda, 2008: Climate Response of Linear and Quadratic Functionals Using the Fluctuation?Dissipation Theorem. *Journal of the Atmospheric Sciences*, **65 (9)**, 2824–2841, doi:10.1175/2007JAS2496.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JAS2496.1.

Hall, A., and S. Manabe, 1997: Can local linear stochastic theory explain sea surface temperature and salinity variability? *Climate Dynamics*, **13 (3)**, 167–180, doi:10.1007/s003820050158, URL http://link.springer.com/10.1007/s003820050158.

Huybers, P., K. A. McKinnon, A. Rhines, and M. Tingley, 2014: U.S. Daily Temperatures: The Meaning of Extremes in the Context of Nonnormality. *Journal of Climate*, **27 (19)**, 7368–7384, doi:10.1175/JCLI-D-14-00216.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-14-00216.1.

Ito, K., 1951: On stochastic differential equations. *Memoirs of the American Mathematical Society*, **0 (4)**, 0–0, doi:10.1090/memo/0004, URL http://www.ams.org/memo/0004.

Jin, F.-F., L. Lin, A. Timmermann, and J. Zhao, 2007: Ensemble-mean dynamics of the ENSO recharge oscillator under state-dependent stochastic forcing. *Geophysical Research Letters*, **34 (3)**, L03 807, doi:10.1029/2006GL027372, URL http://doi.wiley.com/10.1029/2006GL027372.

Johnson, S. D., D. S. Battisti, and E. S. Sarachik, 2000: Empirically Derived Markov Models and Prediction of Tropical Pacific Sea Surface Temperature Anomalies*. *http://dx.doi.org/10.1175/1520-0442(2000)013<0003:EDMMAP>2.0.CO;2*, doi:10.1175/1520-0442(2000)013⟨0003:EDMMAP⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/citedby/10.1175/1520-0442{\%}282000{\%}29013{\%}3C0003{\%}3AEDMMAP{\%}3E2.0.CO{\%}3B2.

Kim, K.-Y., and G. R. North, 1992: Seasonal cycle and second-moment statistics of a simple coupled climate system. *Journal of Geophysical Research*, **97 (D18)**, 20 437, doi:10.1029/92JD02281, URL http://doi.wiley.com/10.1029/92JD02281.

Kimoto, M., and M. Ghil, 1993: Multiple Flow Regimes in the Northern Hemisphere Winter. Part I: Methodology and Hemispheric Regimes. *Journal of the Atmospheric Sciences*, **50 (16)**, 2625–2644, doi:10.1175/1520-0469(1993)050⟨2625:MFRITN⟩2.0.CO;

2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281993{\%}29050{\%}3C2625{\%}3AMFRITN{\%}3E2.0.CO{\%}3B2.

Kolmogoroff, A., 1931: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, **104 (1)**, 415–458, doi:10.1007/BF01457949, URL http://link.springer.com/10.1007/BF01457949.

Kondrashov, D., S. Kravtsov, and M. Ghil, 2006: Empirical Mode Reduction in a Model of Extratropical Low-Frequency Variability. *Journal of the Atmospheric Sciences*, **63 (7)**, 1859–1877, doi:10.1175/JAS3719.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JAS3719.1.

Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability. *Journal of Climate*, **18 (21)**, 4404–4424, doi:10.1175/JCLI3544.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI3544.1.

Leith, C. E., 1975: Climate Response and Fluctuation Dissipation. *Journal of the Atmospheric Sciences*, **32 (10)**, 2022–2026, doi:10.1175/1520-0469(1975)032⟨2022:CRAFD⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281975{\%}29032{\%}3C2022{\%}3ACRAFD{\%}3E2.0.CO{\%}3B2.

Levine, A. F. Z., and F. F. Jin, 2017: A simple approach to quantifying the noise?ENSO interaction. Part I: deducing the state-dependency of the windstress forcing using monthly mean data. *Climate Dynamics*, **48 (1-2)**, 1–18, doi:10.1007/s00382-015-2748-1, URL http://link.springer.com/10.1007/s00382-015-2748-1.

Loikith, P. C., B. R. Lintner, J. Kim, H. Lee, J. D. Neelin, and D. E. Waliser, 2013: Classifying reanalysis surface temperature probability density functions (PDFs) over North America with cluster analysis. *Geophysical Research Letters*, **40 (14)**, 3710–3714, doi:10.1002/grl.50688, URL http://doi.wiley.com/10.1002/grl.50688.

Loikith, P. C., and J. D. Neelin, 2015: Short-tailed temperature distributions over North America and implications for future changes in extremes. *Geophysical Research Letters*, **42 (20)**, 8577–8585, doi:10.1002/2015GL065602, URL http://doi.wiley.com/10.1002/2015GL065602.

Luxford, F., and T. Woollings, 2012: A Simple Kinematic Source of Skewness in Atmospheric Flow Fields. *Journal of the Atmospheric Sciences*, **69 (2)**, 578–590, doi:10.1175/JAS-D-11-089.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JAS-D-11-089.1.

Majda, A. J., C. Franzke, and D. Crommelin, 2009: Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences of the United States of America*, **106 (10)**, 3649–53, doi:10.1073/pnas.0900173106, URL http://www.ncbi.nlm.nih.gov/pubmed/19228943http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2645348.

Majda, A. J., C. Franzke, and B. Khouider, 2008: An applied mathematics perspective on stochastic modelling for climate. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **366 (1875)**, 2427–2453, doi:10.1098/rsta.2008.0012, URL http://www.ncbi.nlm.nih.gov/pubmed/18445572http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2008.0012.

Martinez-Villalobos, C., 2016: Deterministic and stochastic models of tropical climate variability — Cristian Martinez-Villalobos. Ph.D. thesis, University of Wisconsin - Madison, URL https://dept.atmos.ucla.edu/cristian{_}martinez{_}villalobos/publications/deterministic-and-stochastic-models-tropical-climate.

Martinez-Villalobos, C., and D. J. Vimont, 2017: An Analytical Framework for Understanding Tropical Meridional Modes. *Journal of Climate*, **30 (9)**, 3303–3323, doi:10.1175/JCLI-D-16-0450.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0450.1.

Monahan, A. H., 2004: A Simple Model for the Skewness of Global Sea Surface Winds. *Journal of the Atmospheric Sciences*, **61 (16)**, 2037–2049, doi:10.1175/1520-0469(2004)061⟨2037:ASMFTS⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}282004{\%}29061{\%}3C2037{\%}3AASMFTS{\%}3E2.0.CO{\%}3B2.

Moore, A. M., and R. Kleeman, 1999: The Nonnormal Nature of El Ni?o and Intraseasonal Variability. *Journal of Climate*, **12 (10)**, 2965–2982, doi:10.1175/1520-0442(1999)012⟨2965:TNNOEN⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}281999{\%}29012{\%}3C2965{\%}3ATNNOEN{\%}3E2.0.CO{\%}3B2.

Müller, D., 1987: Bispectra of Sea-Surface Temperature Anomalies. *Journal of Physical Oceanography*, **17 (1)**, 26–36, doi:10.1175/1520-0485(1987)017⟨0026:BOSSTA⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0485{\%}281987{\%}29017{\%}3C0026{\%}3ABOSSTA{\%}3E2.0.CO{\%}3B2.

Neelin, J. D., B. R. Lintner, B. Tian, Q. Li, L. Zhang, P. K. Patra, M. T. Chahine, and S. N. Stechmann, 2010: Long tails in deep columns of natural and anthropogenic tropospheric tracers. *Geophysical Research Letters*, **37 (5)**, n/a–n/a, doi:10.1029/2009GL041726, URL http://doi.wiley.com/10.1029/2009GL041726.

Neelin, J. D., S. Sahany, S. N. Stechmann, and D. N. Bernstein, 2017: Global warming precipitation accumulation increases above the current-climate cutoff scale. *Proceedings of the National Academy of Sciences of the United States of America*, **114 (6)**, 1258–1263, doi:

10.1073/pnas.1615333114, URL http://www.ncbi.nlm.nih.gov/pubmed/28115693http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5307437.

Neelin, J. D., and W. Weng, 1999: Analytical Prototypes for OceanAtmosphere Interaction at Midlatitudes. Part I: Coupled Feedbacks as a Sea Surface Temperature Dependent Stochastic Process*. *Journal of Climate*, **12 (3)**, 697–721, doi:10.1175/1520-0442(1999)012⟨0697:APFOAI⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}281999{\%}29012{\%}3C0697{\%}3AAPFOAI{\%}3E2.0.CO{\%}3B2.

Newman, M., 2013: An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies. *Journal of Climate*, **26 (14)**, 5260–5269, doi:10.1175/JCLI-D-12-00590.1.

Newman, M., M. A. Alexander, and J. D. Scott, 2011: An empirical model of tropical ocean dynamics. *Climate Dynamics*, **37 (9-10)**, 1823–1841, doi:10.1007/s00382-011-1034-0, URL http://link.springer.com/10.1007/s00382-011-1034-0.

Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*, doi:10.1002/2017GL074088, URL http://doi.wiley.com/10.1002/2017GL074088.

Newman, M., and Coauthors, 2016: The Pacific Decadal Oscillation, Revisited. *Journal of Climate*, **29 (12)**, 4399–4427, doi:10.1175/JCLI-D-15-0508.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0508.1.

North, G. R., and R. F. Cahalan, 1981: Predictability in a Solvable Stochastic Climate Model. *Journal of the Atmospheric Sciences*, **38 (3)**, 504–513, doi:10.1175/1520-0469(1981)038⟨0504:PIASSC⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281981{\%}29038{\%}3C0504{\%}3APIASSC{\%}3E2.0.CO{\%}3B2.

Papanicolaou, G. C., and W. Kohler, 1974: Asymptotic theory of mixing stochastic ordinary differential equations. *Communications on Pure and Applied Mathematics*, **27 (5)**, 641–668, doi:10.1002/cpa.3160270503, URL http://doi.wiley.com/10.1002/cpa.3160270503.

Peavoy, D., C. L. Franzke, and G. O. Roberts, 2015: Systematic physics constrained parameter estimation of stochastic differential equations. *Computational Statistics & Data Analysis*, **83**, 182–199, doi:10.1016/J.CSDA.2014.10.011, URL https://www.sciencedirect.com/science/article/pii/S016794731400303X.

Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, **98 (2-4)**, 534–558, doi:10.1016/0167-2789(96)00124-8, URL http://linkinghub.elsevier.com/retrieve/pii/0167278996001248.

Penland, C., 2003: A Stochastic Approach to Nonlinear Dynamics: A Review (extended version of the article - ”Noise Out of Chaos and Why it Won’t Go Away”). *Bulletin of the American Meteorological Society*, **84 (7)**, 925–925, doi:10.1175/BAMS-84-7-Penland, URL http://journals.ametsoc.org/doi/abs/10.1175/BAMS-84-7-Penland.

Penland, C., 2007: Stochastic linear models of nonlinear geosystems. *Nonlinear Dynamics in Geosciences*, J. E. A.A. Tsonis, Ed., Springer, New York, chap. 27, 485–515, URL https://www.esrl.noaa.gov/psd/people/cecile.penland/Penland2007.pdf.

Penland, C., and L. Matrosova, 1994: A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation. *Journal of Climate*, **7 (9)**, 1352–1372, doi:10.1175/1520-0442(1994)007⟨1352:ABCFSN⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}281994{\%}29007{\%}3C1352{\%}3AABCFSN{\%}3E2.0.CO{\%}3B2.

Penland, C., and L. Matrosova, 1998: Prediction of Tropical Atlantic Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, **11 (3)**, 483–496, doi:10.1175/1520-0442(1998)011⟨0483:POTASS⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}281998{\%}29011{\%}3C0483{\%}3APOTASS{\%}3E2.0.CO{\%}3B2.

Penland, C., and P. D. Sardeshmukh, 1995: The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, **8 (8)**, 1999–2024, doi:10.1175/1520-0442(1995)008⟨1999:TOGOTS⟩2.0.CO;2, URL http://adsabs.harvard.edu/abs/1995JCli....8.1999P.

Penland, C., and P. D. Sardeshmukh, 2012: Alternative interpretations of power-law distributions found in nature. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **22 (2)**, 023 119, doi:10.1063/1.4706504, URL http://scitation.aip.org/content/aip/journal/chaos/22/2/10.1063/1.4706504.

Perez, C. L., A. M. Moore, J. Zavala-Garay, and R. Kleeman, 2005: A Comparison of the Influence of Additive and Multiplicative Stochastic Forcing on a Coupled Model of ENSO. *Journal of Climate*, **18 (23)**, 5066–5085, doi:10.1175/JCLI3596.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI3596.1.

Perron, M., and P. Sura, 2013: Climatology of Non-Gaussian Atmospheric Statistics. *Journal of Climate*, **26 (3)**, 1063–1083, doi:10.1175/JCLI-D-11-00504.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00504.1.

Rennert, K. J., and J. M. Wallace, 2009: Cross-Frequency Coupling, Skewness, and Blocking in the Northern Hemisphere Winter Circulation. *Journal of Climate*, **22 (21)**, 5650–5666, doi:10.1175/2009JCLI2669.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2009JCLI2669.1.

Ruff, T. W., and J. D. Neelin, 2012: Long tails in regional surface temperature probability distributions with implications for extremes under global warming. *Geophysical Research Letters*, **39 (4)**, n/a–n/a, doi:10.1029/2011GL050610, URL http://doi.wiley.com/10.1029/2011GL050610.

Rümelin, W., 1982: Numerical Treatment of Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, **19 (3)**, 604–613, doi:10.1137/0719041, URL http://epubs.siam.org/doi/abs/10.1137/0719041.

Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2015: Need for Caution in Interpreting Extreme Weather Statistics. *Journal of Climate*, **28 (23)**, 9166–9187, doi:10.1175/JCLI-D-15-0020.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0020.1.

Sardeshmukh, P. D., and C. Penland, 2015: Understanding the distinctively skewed and heavy tailed character of atmospheric and oceanic probability distributions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **25 (3)**, 036 410, doi:10.1063/1.4914169, URL http://scitation.aip.org/content/aip/journal/chaos/25/3/10.1063/1.4914169.

Sardeshmukh, P. D., and P. Sura, 2009: Reconciling Non-Gaussian Climate Statistics with Linear Dynamics. *Journal of Climate*, **22 (5)**, 1193–1207, doi:10.1175/2008JCLI2358.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2008JCLI2358.1.

Sévellec, F., and A. V. Fedorov, 2017: Predictability and Decadal Variability of the North Atlantic Ocean State Evaluated from a Realistic Ocean Model. *Journal of Climate*, **30 (2)**, 477–498, doi:10.1175/JCLI-D-16-0323.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0323.1.

Siegert, S., R. Friedrich, and J. Peinke, 1998: Analysis of data sets of stochastic systems. *Physics Letters A*, **243 (5-6)**, 275–280, doi:10.1016/S0375-9601(98)00283-7, URL https://

www.sciencedirect.com/science/article/pii/S0375960198002837?via{\%}3Dihub.

Smirnov, D., M. Newman, and M. A. Alexander, 2014: Investigating the Role of OceanAtmosphere Coupling in the North Pacific Ocean. *Journal of Climate*, **27 (2)**, 592–606, doi:10.1175/JCLI-D-13-00123.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-13-00123.1.

Smyth, P., K. Ide, and M. Ghil, 1999: Multiple Regimes in Northern Hemisphere Height Fields via MixtureModel Clustering*. *Journal of the Atmospheric Sciences*, **56 (21)**, 3704–3723, doi:10.1175/1520-0469(1999)056⟨3704:MRINHH⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469{\%}281999{\%}29056{\%}3C3704{\%}3AMRINHH{\%}3E2.0.CO{\%}3B2.

Stechmann, S. N., and J. D. Neelin, 2014: First-Passage-Time Prototypes for Precipitation Statistics. *Journal of the Atmospheric Sciences*, **71 (9)**, 3269–3291, doi:10.1175/JAS-D-13-0268.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JAS-D-13-0268.1.

Stefanova, L., P. Sura, and M. Griffin, 2013: Quantifying the Non-Gaussianity of Wintertime Daily Maximum and Minimum Temperatures in the Southeast. *Journal of Climate*, **26 (3)**, 838–850, doi:10.1175/JCLI-D-12-00161.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-12-00161.1.

Stratonovich, R. L., 1966: A New Representation for Stochastic Integrals and Equations. *SIAM Journal on Control*, **4 (2)**, 362–371, doi:10.1137/0304028, URL http://epubs.siam.org/doi/10.1137/0304028.

Sura, P., and A. Hannachi, 2015: Perspectives of Non-Gaussianity in Atmospheric Synoptic and Low-Frequency Variability. *Journal of Climate*, **28 (13)**, 5091–5114, doi:10.1175/JCLI-D-14-00572.1, URL http://journals.ametsoc.org/doi/10.1175/JCLI-D-14-00572.1.

Sura, P., and M. Newman, 2008: The Impact of Rapid Wind Variability upon AirSea Thermal Coupling. *Journal of Climate*, **21 (4)**, 621–637, doi:10.1175/2007JCLI1708.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1708.1.

Sura, P., M. Newman, and M. A. Alexander, 2006: Daily to Decadal Sea Surface Temperature Variability Driven by State-Dependent Stochastic Heat Fluxes. *Journal of Physical Oceanography*, **36 (10)**, 1940–1958, doi:10.1175/JPO2948.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JPO2948.1.

Sura, P., M. Newman, C. Penland, and P. Sardeshmukh, 2005: Multiplicative Noise and Non-Gaussianity: A Paradigm for Atmospheric Regimes? *Journal of the Atmospheric Sciences*, **62 (5)**, 1391–1409, doi:10.1175/JAS3408.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JAS3408.1.

Sura, P., and P. D. Sardeshmukh, 2008: A Global View of Non-Gaussian SST Variability. *Journal of Physical Oceanography*, **38 (3)**, 639–647, doi:10.1175/2007JPO3761.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JPO3761.1.

Sura, P., and P. D. Sardeshmukh, 2009: A global view of airsea thermal coupling and related non-Gaussian SST variability. *Atmospheric Research*, **94 (1)**, 140–149, doi:10.1016/j.atmosres.2008.08.008, URL http://linkinghub.elsevier.com/retrieve/pii/S0169809508002391.

Thompson, C. J., and D. S. Battisti, 2000: A Linear Stochastic Dynamical Model of ENSO. Part I: Model Development. *Journal of Climate*, **13 (15)**, 2818–2832, doi:10.1175/1520-0442(2000)013⟨2818:ALSDMO⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0442{\%}282000{\%}29013{\%}3C2818{\%}3AALSDMO{\%}3E2.0.CO{\%}3B2.

Timmermann, A., H. U. Voss, and R. Pasmanter, 2001: Empirical Dynamical System Modeling of ENSO Using Nonlinear Inverse Techniques. *Journal of Physical Oceanography*, **31 (6)**, 1579–1598, doi:10.1175/1520-0485(2001)031⟨1579:EDSMOE⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0485{\%}282001{\%}29031{\%}3C1579{\%}3AEDSMOE{\%}3E2.0.CO{\%}3B2.

Vimont, D. J., 2010: Transient Growth of Thermodynamically Coupled Variations in the Tropics under an Equatorially Symmetric Mean State*. *Journal of Climate*, **23 (21)**, 5771–5789, doi:10.1175/2010JCLI3532.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3532.1.

Vimont, D. J., 2012: Analysis of the Atlantic Meridional Mode Using Linear Inverse Modeling: Seasonality and Regional Influences. *Journal of Climate*, **25 (4)**, 1194–1212, doi:10.1175/JCLI-D-11-00012.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00012.1.

Vimont, D. J., M. A. Alexander, and M. Newman, 2014: Optimal growth of Central and East Pacific ENSO events. *Geophysical Research Letters*, **41 (11)**, 4027–4034, doi:10.1002/2014GL059997, URL http://doi.wiley.com/10.1002/2014GL059997.

von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, 1988: Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere. *Journal of Geophysical Research*, **93 (D9)**, 11 022, doi:10.1029/JD093iD09p11022, URL http://doi.wiley.com/10.1029/JD093iD09p11022.

Wilkins, J. E., 1944: A Note on Skewness and Kurtosis. *The Annals of Mathematical Statistics*, **15 (3)**, 333–335, doi:10.1214/aoms/1177731243, URL http://projecteuclid.org/euclid.aoms/1177731243.

Wu, R., B. P. Kirtman, and K. Pegion, 2006: Local AirSea Relationship in Observations and Model Simulations. *Journal of Climate*, **19 (19)**, 4914–4932, doi:10.1175/JCLI3904.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI3904.1.

Zanna, L., 2012: Forecast Skill and Predictability of Observed Atlantic Sea Surface Temperatures. *Journal of Climate*, **25 (14)**, 5047–5056, doi:10.1175/JCLI-D-11-00539.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00539.1.

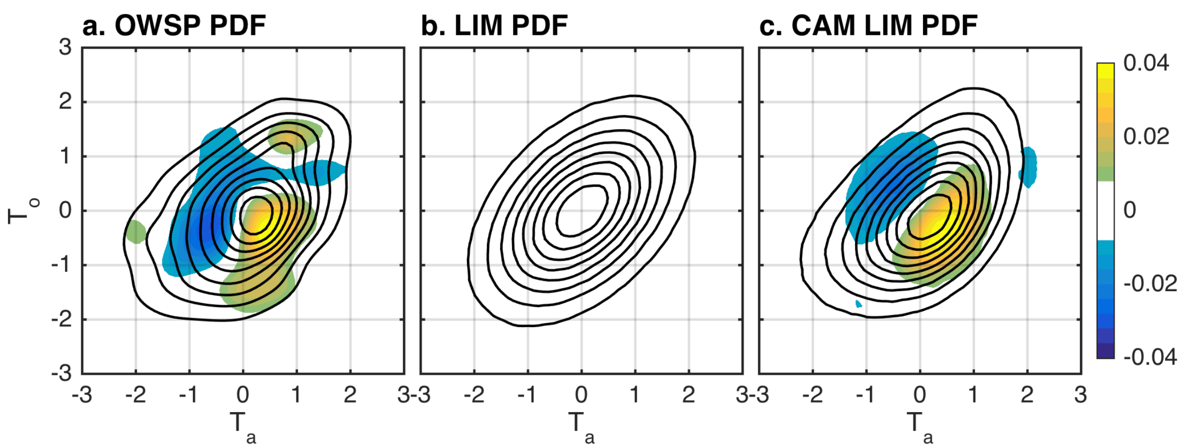
Zanna, L., and E. Tziperman, 2005: Nonnormal Amplification of the Thermohaline Circulation. *Journal of Physical Oceanography*, **35 (9)**, 1593–1605, doi:10.1175/JPO2777.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JPO2777.1.

LIST OF TABLES

TABLE 1. Observed and modeled skewness and kurtosis.

Variable	Skewness	Kurtosis
T_a (obs)	-0.51	3.78
T_a (CAM-LIM)	-0.55	3.80
T_a (LIM)	0	3.00
T_o (obs)	0.51	3.94
T_o (CAM-LIM)	0.41	3.61
T_o (LIM)	0.01	3.00

LIST OF FIGURES



976 FIG. 1. Three day running mean T_o and T_a joint PDFs (solid), calculated using **a** Observed data 1950-1980
977 November - April, **b** LIM full integration, **c** CAM-LIM full integration. Shading denotes differences from a best
978 fit bivariate Gaussian distribution. Units are of standard deviation.

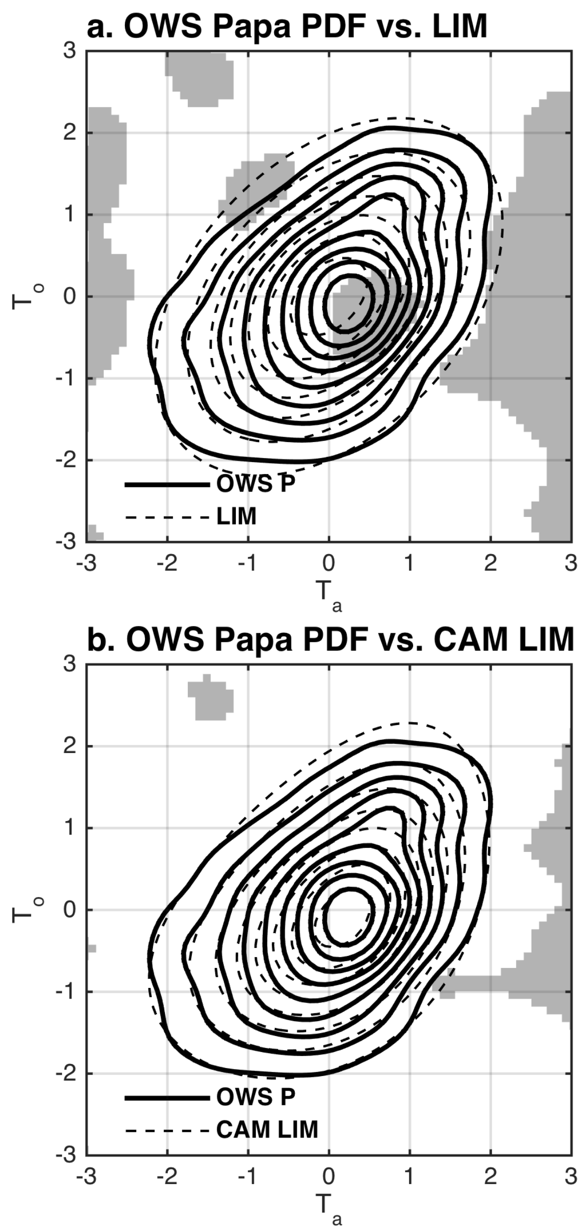


FIG. 2. Comparison of T_a, T_o observed joint PDF (solid) and (**a**) LIM and (**b**) CAM-LIM generated joint PDFs of the same length as the observations (617 realizations). The dashed line denotes the average of the 617 LIM and CAM-LIM realizations, and shading denotes region where the observed joint PDF is outside the 2.5 and 97.5 percentile estimated from the LIM and CAM-LIM realizations.

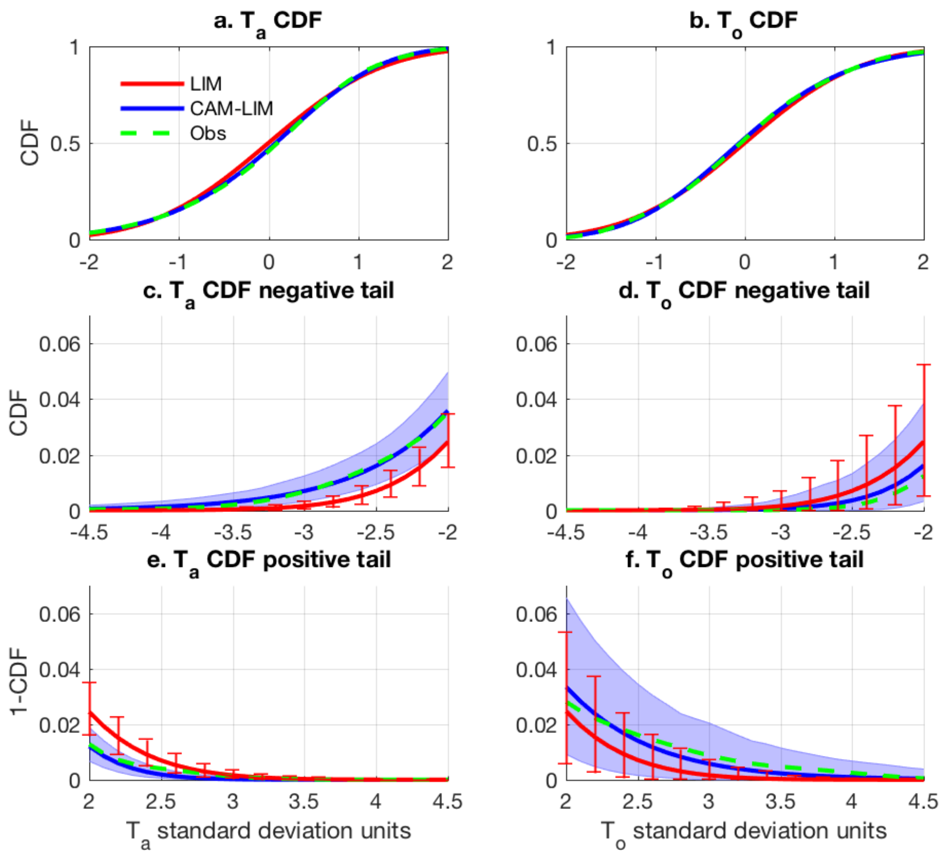


FIG. 3. T_a (left panel, **a**, **c**, **e**) and T_o (right panel, **b**, **d**, **f**) LIM generated (red), CAM-LIM generated (blue), and observed (green dashed) cumulative density functions. Solid red and blue lines denote the average of 617 different LIM and CAM-LIM realizations the same length as the observations, with confidence intervals showing the region within the 2.5 percentile and 97.5 percentile of the realizations (LIM red error bars, CAM-LIM blue shading). Units are of standard deviation. Top panel (**a**, **b**) shows the middle range of the data (between -2 and 2 standard deviations), middle panel (**c**, **d**) shows the negative tail, and the bottom panel (**e**, **f**) the positive tail. All CDFs are estimated using an Epanechnikov kernel (Epanechnikov 1969; Bowman and Azzalini 1997). Figures c-f are also shown in a logarithmic y axis in figure S3.

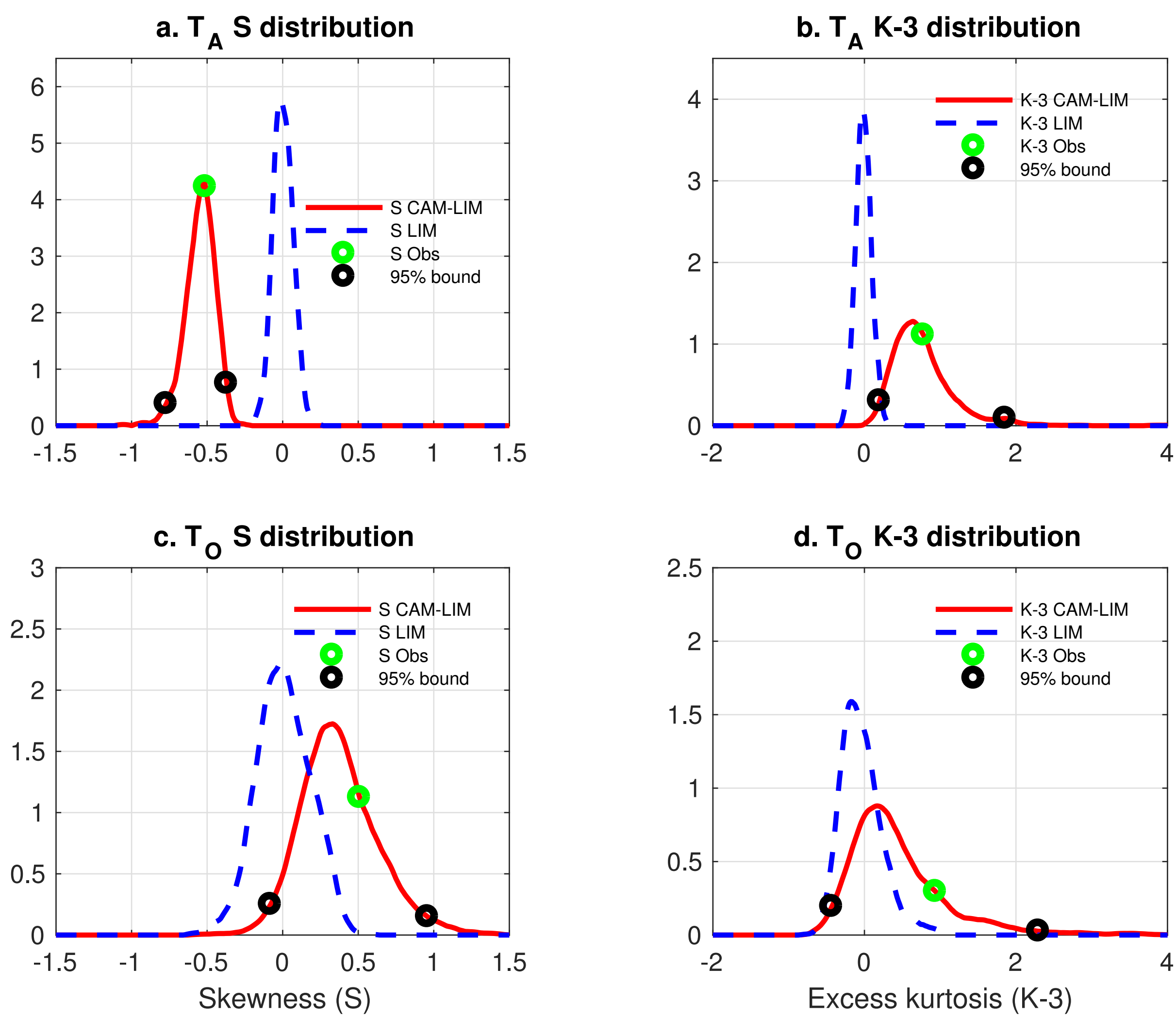


FIG. 4. An estimation of (T_a, T_o) skewness (**a**, **c**), and (T_a, T_o) excess kurtosis (**b**, **d**) distributions using an Epanechnikov kernel. This is calculated by dividing the full LIM (blue solid), and CAM-LIM (red dashed) generated datasets into segments the length of the observed dataset (617 realizations). In each subplot the light green circle denotes the observed skewness or excess kurtosis, and the black circles denote the 2.5 and 97.5 percentile bounds. The observed values are within the 95% confidence level generated by the CAM-LIM and outside the confidence level generated by the LIM in all cases.

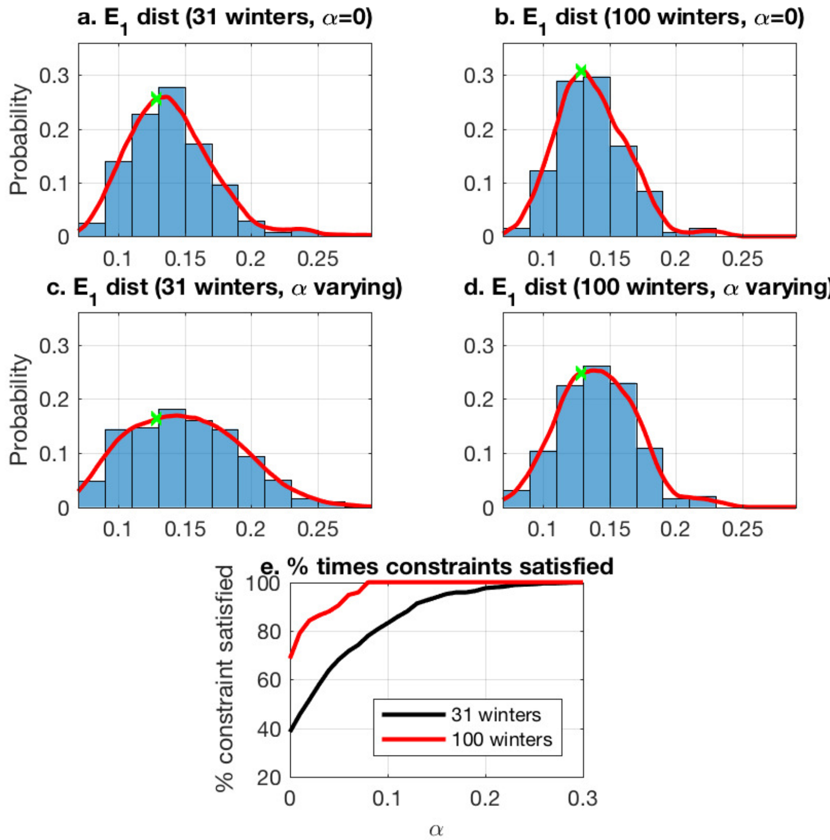


FIG. 5. A histogram (with an Epanechnikov kernel estimated PDF superimposed) of E_1 retrieved values calculated using segments of (**a**, **c**) 31 winters length, and (**b**, **d**) segments of 100 winters length of the full CAM-LIM generated dataset. Here we show two different cases. (**a**, **b**) Shows the histogram of the retrieved values using segments where constraints (24) and (25) are satisfied ($\alpha = 0$), and (**c**, **d**) Shows retrieved values for all segments, after the methodology outlined in section 4c is followed (α varying). In this case for each segment we use a redefined matrix $(1+\alpha)\mathbf{K}$, with $\alpha > 0$ increasing from 0 (with increments of 0.01) up to the point where (24) and (25) are satisfied. The histogram represents the probability for a value inside of each bin (of width 0.02) to be retrieved, and the PDF is normalized to match the height of the histogram. **e** Shows the percentage of segments where constraints are satisfied as a function of α, for both 31 and 100 winters segments length.